2

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
NAVAL AIR STATION, PENSACOLA, FLORIDA 32508-5700

AD-A238 564

NAMRL-1361

PERFORMANCE OF SEVERAL DIFFERENT NAVAL AVIATOR COMMUNITIES ON A COGNITIVE/PSYCHOMOTOR TEST BATTERY: PIPELINE COMPARISON AND PREDICTION

R.N. SHULL AND G.R. GRIFFIN

DTIC
ELECTE
JUL 23 1991
S B D

DEFENSE TECHNICAL INFORMATION CENTER
9105535

This document has been approved for public release; distribution is unlimited.

91 7 18 027

Reviewed and approved 31 December 1990

J. A. BRADY, CAPT, MSC USN
Commanding Officer

The research reported in this paper was completed under the Naval Medical Research and Development Command work unit 63706N M0096 M00960.01 0151.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Trade names of materials and/or products of commercial or nongovernment organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products.

Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines.

Reproduction in whole or in part is permitted for any purpose of the United States government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1990	Interim Dec 88 - Dec 89

4. TITLE AND SUBTITLE
Performance of Several Different Naval Aviator Communities on a Cognitive/Psychomotor Test Battery: Pipeline Comparison and Prediction

5. FUNDING NUMBERS
63706N M0096 M00960.01

6. AUTHOR(S)
R. N. Shull and G. R. Griffin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, FL 32508-5700

8. PERFORMING ORGANIZATION REPORT NUMBER
NAMRL-1361

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center, Bldg. 1
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Research is currently being conducted to develop reliable predictor tests which might aid in the making of decisions concerning aircrew assignment. The current approach involves comparing the performance of several different aviator communities on a test battery measuring various aspects of cognitive and psychomotor functioning. Four groups of pilots (two experienced jet groups, one experienced helicopter (helo) group, and one untrained student group) were tested on this battery. Overall, the jet groups performed in an equivalent manner, while the helo group showed a lower test performance level in comparison. The student group performed at a lower level than the experienced groups in general. Within this student group, pilot trainees who were assigned to the jet pipeline did significantly better on several of these tests than those trainees who were assigned to either helicopter or land-based fixed-wing pipelines. Many of the test performance differences seen between these jet and helo student pipeline groups were also seen between the experienced jet and helo pilots tested.

14. SUBJECT TERMS
Pilot Selection, Pipeline Assignment, Psychomotor Ability, Cognitive Ability, Jet Pilots, Helicopter Pilots, Student Pilots, Multitask Performance, Longterm/Innate Ability Measures

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18

SUMMARY PAGE

THE PROBLEM

Several studies have suggested the possibility of predicting an aviator's operational performance in various training and fleet aviation environments. Research is currently being conducted to develop reliable predictor tests that might aid in decisions concerning aircrew selection, training pipeline assignment, and posttraining aircraft assignment. We compared the performance of four groups of pilots (two experienced jet groups, one experienced helicopter [helo] group, and one untrained student group) on a test battery measuring various aspects of cognitive and psychomotor function.

FINDINGS

Overall, the jet groups performed similarly to each other and better than the helo group. The student group performed at a lower level than the experienced groups in general. Within the student group, pilot trainees assigned to the jet pipeline did significantly better on several tests than those assigned to either the helicopter or land-based fixed-wing pipelines. Many of the test performance differences seen between these jet and helo student pipeline groups were also seen between the experienced jet and helo pilots tested. Perhaps these tests are measuring an innate ability that remains relatively constant throughout a pilot's career in comparison to other pilots with equivalent levels of flight experience.

RECOMMENDATIONS

Research with this test battery should be continued to ascertain differences in test performance among other pilot groups not yet examined. An analysis of the relationship between test performance and other phases of training, as well as the interaction of these tests with other available aviation selection criteria, is warranted. Finally, human factors task analysis might be useful in determining appropriate selection criteria for a particular aircraft type.

Acknowledgments

The authors gratefully acknowledge Mr. Peter Collyer and Ms. I. Catherine Vogel for their assistance in programming and modifying the cognitive/psychomotor test battery and Mr. Alfred Thomas and Mr. Soloman Eagles for their care and diligence in testing subjects.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

INTRODUCTION

Research is being performed at the Naval Aerospace Medical Research Laboratory (NAMRL) to develop measures of cognitive and psychomotor ability that would demonstrate a reliable relationship to the flight performance of naval aviators. The goal of such efforts is the eventual development of a test battery that would predict the operational performance of future naval aviators. A test battery of this type would aid in the identification of unique selection criteria for specific fleet aviation communities which, in turn, would support platform assignment (pipeline) decisions. Thus, the Navy's capability to assign aviators to flight environments in which they would be most effective could be increased with a commensurate improvement in operational readiness and safety.

A number of research efforts have proven somewhat successful in the prediction of certain measures of operational pilot performance. For example, peer ratings obtained during preflight training were useful in identifying both successful and unsuccessful naval aviators during combat in Vietnam (1). During the midsixties (2), F-4 Replacement Air Group (RAG) training evaluation aided in the development of a prediction equation that had the possibility of reducing RAG attrition by 38%. Successful predictions of F-4 carrier landing performance have been made (3) using a combination of psychological tests and actual flight performance measures. Also, a regression equation based on the performance of an East coast F-4 RAG reliably predicted performance of a West coast F-4 RAG, and an overall experience measure, combined with seven undergraduate training grades reliably predicted the overall RAG grade (4,5). Recently (6), a set of automated dichotic listening and psychomotor (cursor tracking) test results appeared to be significantly correlated with some elements of the Air Combat Maneuvering (ACM) performance of a group of Marine F-4 pilots.

These studies suggest the possibility of successfully predicting at least some elements of operational pilot performance in various naval aviation environments. Our approach has been to use an automated battery of cognitive and psychomotor tests to attempt to predict pilot performance in specific operational settings. In previous research using this approach (7,8), we found no significant relationships between performance on these tests and either Fleet Replacement Squadron (FRS) or ACM performance in tactical jet aircraft. One reason may be that the skill and ability levels found within the pilot groups tested (F/A-18 for FRS and F-14 for ACM) were already equalized across their respective members due to common selection, training, and flight experiences. Such research does not address the critical issue of test performance differences among pilots of different aircraft. To be useful in making pipeline decisions, these test measures must reliably distinguish among the different aviation communities.

The first part of this report documents a comparison of the test battery performance among the jet pilot groups mentioned above and a group of helicopter instructor pilots. We also included the test results of a group of student naval aviators (SNAs) in this comparison to illustrate the possible effects of training on test performance. The second part of this report documents a comparison of SNA test performance, before any military flight

training, among the various pipeline groupings and analyzes the relationship of this performance to both flight grade and pipeline assignment.

METHODS

SUBJECTS

We tested four different groups of volunteer male subjects as described in Table 1. Group 1 consisted of Navy F-14 pilots who were participating in the Fleet Fighter ACM Readiness Program against the VF-43 adversary squadron at NAS Oceana, Virginia. Group 2 contained Navy jet pilots assigned to VFA-106 at NAS Cecil Field, Florida, who were completing FRS training for transition to the F/A-18. About half of this group had been assigned to the F/A-18 directly after completing advanced undergraduate flight training while the other half was transitioning from other fleet aircraft, typically A-7s or F-4s. Group 3 consisted of fleet helicopter pilots assigned to either HT-8 or HT-18 at NAS Whiting Field, Florida, who were involved in the initial training of student helicopter pilots preparing for their transition to fleet aircraft. Group 4 consisted of SNAs assigned to Naval Aviation Schools Command, NAS Pensacola, Florida, who were preparing to receive primary flight training. Before testing, all SNAs were informed that test results would not affect their continuation in the flight program and would not be entered into their service record.

TABLE 1. Age (Mean ± SD) and Flight Time (Mean ± SD) of Subjects.

Group	n	Type	Age (years)		Flight time (h)	
1	66	F-14	24 - 41	(29.09 ± 4.11)	350 - 4500	(1473 ± 1068)
2	67	F/A-18	24 - 41	(28.91 ± 3.47)	100 - 3880	(942 ± 859)
3	39	helo	26 - 36	(29.44 ± 2.06)	750 - 4950	(1936 ± 851)
4	200	SNA	20 - 27	(22.91 ± 1.40)	0 - 275	(13 ± 39)

APPARATUS AND PROCEDURES

Table 2 lists the tests given and their sequence. The entire series was automated using an Apple IIe microcomputer, an Amdek Color I Plus monitor (CRT), and an Apple IIe numeric keypad. Subjects received all test instructions on the CRT before each test began. The entire test battery took 90 - 120 min to complete, depending on the number and type of test elements given each subject group. The test types are described in the following sections; further details on each type may be found elsewhere (7).

TABLE 2. Description and Sequence of Automated Tests.

Presentation order	Description
1.	Single psychomotor task (PMT), stick only (S)
2.	Single dichotic listening task (DLT)
3.	First multitask (1,2 combined)
4.	Single (PMT), stick and rudder (SR)
5.	Second multitask (4,2 combined)
6.	Third multitask (4,2 combined)
7.	Single PMT; stick, rudder, and throttle (SRT)
8.	Second single PMT (like 7, SRT)
9.	Fourth multitask (8,2 combined)
10.	One dimensional compensatory tracking (ODCT)
11.	Absolute difference computation (ADC)
12.	Fifth multitask, ODCT and ADC (10,11 combined)

Psychomotor Task (PMT)

The psychomotor tracking task required subjects to maintain first one, then two, and finally three randomly displaced cursors on fixed targets on the CRT by manipulating joysticks and foot pedals. Subjects manipulated one Measurement Systems, Inc., joystick (stick or S) at the front seat edge with their right hand to control a cursor that moved within the upper two-thirds of the screen just right of center in a backwards (reversed) manner. Locally produced rudder pedals (rudder or R) patterned after those of a Systems Research Laboratories, Inc., psychomotor test device were used to control a cursor that moved horizontally across the bottom of the screen. Pushing the left pedal moved this cursor to the right while pushing the right pedal moved it to the left. Another Measurement Systems joystick (throttle or T) located on the left seat edge was manipulated by the subject's left hand to move a cursor vertically on the left side of the screen. The subject pulled this throttle back to move this cursor down and vice versa.

Psychomotor task tests 1, 4, and 7 (see Table 2) were each preceded by a 3-min practice period. For the F/A-18 pilot group, test 1 was divided into two 3-min testing periods separated by a 20-sec rest period while test 4 was divided into three 3-min testing periods separated by two 20-s rest periods. For the other three subject groups, test 4 was divided into two 3-min testing periods separated by a 20-s rest period. Test 1 for these three groups and tests 7 and 8 for all the groups had a single 3-min testing period each. Psychomotor task scores were the accumulated total of absolute deviations from an ideal target position. For each time sampling of cursor position, absolute pixel errors were assessed along each dimension separately. The final error score was the sum of all the samplings made across all the dimensions represented in that particular task. This error score was for the total time of that test and was then divided by the number of minutes of that test, which

yielded a standard rate of pixel error per minute of test time. The scores of tests 5 and 6 and tests 7 and 8 were averaged for each subject. All of these PMT error scores were then transformed by using logarithms to base 10 to reduce skewness and compensate for extreme outliers, thus reducing the complexity of data analysis while retaining all the data points available (11).

Dichotic Listening Task (DLT)

The DLT consisted of a series of letter/digit string sets presented aurally over binaural headphones via two Jameco JE 520-AP voice synthesizers. Subjects were told which ear to attend to for each trial; first for a series of 16 pairs of letters and/or numbers and then again for a series of 6 more pairs. Subjects were to indicate the digits (0-9) presented to the designated ear in the order of their occurence. Subjects responded with their left hand using a separate keypad placed immediately in front and slightly left of center. The test was preceded by six aural practice trials, which provided immediate performance feedback, visually indicating the letters and digits presented and the subjects' keypad responses. Subjects also completed three multiple-choice questions before the start of this test to ensure that they understood the concept of the DLT.

The DLT performance measure for the F/A-18 pilot group and about half of the SNA group was the number of incorrect responses made during 24 trials, in which a total of 216 correct responses were possible, with this number being divided by two. This made the score directly comparable to those of the rest of the subject groups whose performance measure involved only 12 trials in which a total of 108 responses were possible. In all cases, the number of correct responses made was subtracted from 108, and after adding 1.0, this new adjusted error score was then transformed by using logarithms to base 10 to adjust for both skewness and extreme outliers as was done for the PMT (11).

Multitask PMT/DLT

In all of the multitask conditions, subjects performed both the DLT and PMT simultaneously, (a 12-trial DLT and a 4.5-min PMT). During the first multitask condition, subjects performed both the DLT and the PMT(S) as described above. During the next two multitask conditions, subjects performed both the DLT and the PMT(SR) using their right hand and their feet to control the central joystick and the rudder pedals, and their left hand to make keypad responses to the DLT input. During the final multitask condition, subjects performed both the DLT and the PMT(SRT). In this most elaborate combination, subjects used their right hand and both feet to control the central joystick and the rudder pedals as before but, in addition, used their left hand to control the throttle joystick and voiced their DLT responses using a headset microphone. The vocal responses were tape-recorded for subsequent analysis and hand scoring. Before the start of the various multitask combinations, subjects were instructed to perform each task equally well. Performance measures for the PMT and DLT in these multitask conditions were identical to those of the single tasks with PMT errors being recorded for the final 4 min of each test.

One Dimensional Compensatory Tracking (ODCT)

The task required subjects to center a square cursor inside of an elongated rectangle by making left and right movements of a joystick centered on the front seat edge with their right hand. The cursor is driven by a forcing function, which increases centering effort with distance from center. Both jet pilot groups received three 2-min trials, with each trial separated by a 30-s rest period. The ODCT test performance measure for these groups was total pixel deviation error averaged over these three single task trials. The helicopter pilot group received four 2-min trials, separated by the same 30-s rest periods, with total error averaged over the first three trials. The SNA group received ten 2-min trials, separated by these same rest periods, with total error averaged over the last three trials.

Absolute Difference Computation (ADC)

Randomly selected digits between 1 and 9 were presented inside a small square in the middle of the CRT to subjects who then determined the absolute difference between the digit currently displayed on the CRT and the last digit displayed previously. The subjects then pressed the corresponding digit-key on the keypad with their left hand as quickly as possible resulting in the display of another number for computation. Identical digits were not allowed to repeat, and only the digit responses 1, 2, 3, and 4 were possible. Both jet groups received three 2-min trials, with each trial separated by a 20-s rest period. Performance measures for the ADC were the number of correct responses (CR) and incorrect responses (IR) made, both averaged over the three ADC alone trials. Another test performance measure involved the division of the ADC IR measure into the ADC CR measure (ADC CR/IR) in order to determine how many correct responses were made for every incorrect response during either ADC test. The helicopter group received four 2-min trials with the same rest periods and performance measures averaged over the first three trials. The SNA group received fifteen 2-min trials, with the same rest periods and performance measures averaged over the last three trials.

Dual-Task ODCT/ADC

During this phase of testing, subjects performed both the ODCT and ADC concurrently. The digits for the difference task were centered just above the tracking task. The subjects controlled the tracking task joystick with their right hand and made keypad responses to the difference task with their left hand. Subjects were instructed to perform each task equally well. Both jet groups received three 2-min trials with each trial separated by a 30-s rest period. Test measures for the dual task ODCT/ADC were the same as those for the single tasks. The helicopter group received four 2-min trials, with the same rest periods and performance measures averaged over the first three trials. The SNA group received five 2-min trials, with the same rest periods and performance measures averaged over the last three trials.

RESULTS AND DISCUSSION

AVIATOR TEST PERFORMANCE (PART 1)

Table 3 presents descriptive statistics of the test performance of the four subject groups on these cognitive and psychomotor tests. Of the 200 SNAs who were tested, 177 eventually passed primary flight training. To more closely match this SNA group to the others tested (all of whom passed such training), only those who passed were included in the analysis. Due to technical difficulties, the results of both the last DLT test (test 9) for the F-14 group and the incorrect response (IR) measures of the ADC tests for the F/A-18 group were not available for analysis and thus were not included in the table. The SNA group was not tested on the multitask PMT(SRT) w/DLT and thus these scores were also not included in the table.

TABLE 3. Descriptive Statistics [Mean (± SD] of Tests for All Groups.

Test measure	F-14 (n = 66)		F/A-18 (n = 67)		HELO (n = 39)		SNAs (n = 177)	
DLT alone	0.72	(0.34)	0.71	(0.23)	0.85	(0.26)	0.76	(0.26)
DLT w/PMT(S)	0.84	(0.34)	0.65	(0.37)	1.04	(0.30)	0.87	(0.36)
DLT w/PMT(SR)	0.81	(0.24)	0.74	(0.29)	0.92	(0.26)	0.93	(0.30)
DLT w/PMT(SRT)	----	----	0.87	(0.31)	1.09	(0.25)	----	----
PMT(S) alone	3.03	(0.20)	3.03	(0.13)	3.17	(0.32)	3.36	(0.29)
PMT(S) w/DLT	2.79	(0.15)	2.74	(0.15)	2.97	(0.25)	3.09	(0.30)
PMT(SR) alone	3.43	(0.13)	3.39	(0.12)	3.45	(0.18)	3.66	(0.21)
PMT(SR) w/DLT	3.16	(0.14)	3.14	(0.17)	3.23	(0.20)	3.43	(0.24)
PMT(SRT) alone	3.59	(0.13)	3.56	(0.16)	3.60	(0.16)	3.73	(0.18)
PMT(SRT) w/DLT	3.43	(0.19)	3.37	(0.14)	3.45	(0.18)	----	----
Sgl ODCT	19.31	(7.76)	23.08	(6.79)	29.61	(9.04)	31.19	(13.20)
Sgl ADC CR	58.63	(15.63)	56.73	(13.16)	45.60	(16.16)	75.71	(15.96)
Sgl ADC IR	6.15	(5.54)	----	----	7.40	(12.36)	12.47	(28.08)
Sgl ADC CR/IR	18.86	(18.64)	----	----	12.28	(8.45)	25.50	(36.96)
Dual ODCT	29.28	(11.85)	33.90	(11.41)	42.96	(13.43)	36.14	(15.92)
Dual ADC CR	62.68	(15.45)	60.94	(13.32)	53.37	(15.96)	67.46	(16.54)
Dual ADC IR	6.98	(9.20)	----	----	4.99	(2.85)	11.86	(21.89)
Dual ADC CR/IR	25.65	(36.88)	----	----	16.08	(13.57)	12.80	(14.15)

Most of the tests showed a significant difference among subject groups tested using one-way analysis of variance (ANOVA) except for the DLT alone, the PMT(SRT) w/DLT, and the IR measure for both the single and dual ADC test (Table 4).

Using the Scheffe post-hoc comparison test (9), several intergroup pair-wise comparisons were significant at both the .05 and .10 alpha level. Table 5 shows a graphic representation of these comparisons in which the test scores involved are presented in declining order of magnitude for each test.

For the DLT w/PMT(S) measure, the helicopter (helo) group made more errors than the other groups while the F/A-18 group made fewer errors than the other groups. For the DLT w/PMT(SR) measure, the F/A-18 group made fewer errors than either the helo or SNA groups. For the DLT w/PMT(SRT) measure, the F/A-18 group made fewer errors than the helo group. For the PMT(S) alone and w/DLT measures, the SNA group made more errors than any of the other groups while the helo group made more errors than either jet group. For the PMT(SR) alone and w/DLT measures, as well as for the PMT(SRT) alone measure, the SNA group made more errors than any of the other groups.

For the single ODCT measure, the jet groups made fewer errors than either the SNA or helo groups. For the single ADC CR measure, the SNA group made more correct responses than any of the other groups while the helo group made more correct responses than either jet group. For the dual ODCT measure, the helo group made more errors than either jet group while for the dual ADC CR measure, the helo group made fewer correct responses than either the F-14 or SNA group. For the single ADC CR/IR measure, the SNA group did better than the helo group while, for the dual ADC CR/IR measure, the F-14 group did better than the SNA group. The SNAs were also the only group to show a significant change (decrease) in this measure between the single and dual ADC tests (t = 5.20, n = 174, p < .01). Only the helo group had a significant proportion of Marine Corps subjects, and no significant differences in test performance were found between those Marine Corps helo pilots tested (n = 10) and those of the Navy (n = 29).

TABLE 4. Analysis of Variance of Tests for All Groups Tested.

Test measure	F	df	p
DLT alone	2.57	3, 343	> .05
DLT w/PMT(S)	10.76	3, 343	< .0001
DLT w/PMT(SR)	8.37	3, 343	< .0001
DLT w/PMT(SRT)	11.56	1, 72	< .002
PMT(S) alone	41.83	3, 342	< .0001
PMT(S) w/DLT	45.02	3, 345	< .0001
PMT(SR) alone	54.28	3, 340	< .0001
PMT(SR) w/DLT	48.96	3, 344	< .0001
PMT(SRT) alone	20.19	3, 315	< .0001
PMT(SRT) w/DLT	2.29	2, 137	> .10
Sgl ODCT	20.52	3, 315	< .0001
Sgl ADC CR	52.16	3, 310	< .0001
Sgl ADC IR	1.34	2, 266	> .20
Sgl ADC CR/IR	3.11	2, 266	< .05
Dual ODCT	7.56	3, 315	< .0002
Dual ADC CR	8.98	3, 315	< .0001
Dual ADC IR	1.47	2, 275	> .20
Dual ADC CR/IR	4.53	2, 275	< .02

Interestingly, for all subject groups, the mean number of errors made on the PMT decreased when the DLT was added, regardless of motor complexity level. Two-tailed t tests for dependent samples showed this difference to be significant for all conditions (all t values > 4.08, all n values > 38, all p values < .01) and would indicate that all the subject groups performed better on the PMT when it was combined with the DLT. In fact, as the DLT was brought on-line with the PMT, the particular microcomputer used could not maintain the level of cursor positioning difficulty attained previously due to processor overload. This overloading also produced a possible reduction in error sampling rate as test complexity increased. This apparent decrease in testing efficiency does not invalidate the usefulness of these results or methodology. Using Friedman two-way ANOVAs (10), we found that for all the subject groups, significantly more errors were made as PMT complexity increased during both the unitask and multitask conditions (all ANOVA chi-square values > 24.64, all p values < .01).

Table 5. Intergroup Comparisons for Scores from Each Test (declining magnitude represents improved performance for all test measures except ADC CR and ADC CR/IR).

Test measure	Significant intergroup differences
DLT alone	No significant differences
DLT w/PMT(S)	HELO > SNAs, F-14 > F/A-18
DLT w/PMT(SR)	SNAs, HELO > F/A-18
DLT w/PMT(SRT)	HELO > F/A-18
PMT(S) alone	SNAs > HELO > F-14, F/A-18
PMT(S) w/DLT	SNAs > HELO > F-14, F/A-18
PMT(SR) alone	SNAs > HELO, F-14, F/A-18
PMT(SR) w/DLT	SNAs > HELO, F-14, F/A-18
PMT(SRT) alone	SNAs > HELO, F-14, F/A-18
PMT(SRT) w/DLT	No significant differences
Sgl ODCT	SNAs, HELO > F/A-18, F-14
Sgl ADC CR	SNAs > F-14, F/A-18 > HELO
Sgl ADC IR	No significant differences
Sgl ADC CR/IR	SNAs > HELO
Dual ODCT	HELO > F/A-18, F-14
Dual ADC CR	SNAs, F-14 > HELO
Dual ADC IR	No significant differences
Dual ADC CR/IR	F-14 > SNAs

SNA SUBGROUP TEST PERFORMANCE (PART 2)

Table 6 presents descriptive statistics of test performance of the various pipeline subgroupings that made up the SNA subject group after they passed primary flight training. The test labeling is the same as that used in Table 3 and the subgroups are labelled according to platform assignment. The term JETS refers to any jet powered tactical aircraft, PROP refers to

land-based propellor driven fixed-wing aircraft, HELO refers to all rotary wing aircraft, and E2C2 refers to a special class of carrier-based propellor driven fixed wing aircraft. This table also includes descriptive statistics on the age, final primary flight grade, and number of hours in the training aircraft needed to complete primary training for each SNA subgroup.

TABLE 6. Descriptive Statistics of Tests for SNA/Pipeline Groups [Mean (± SD)].

Test measure	JETS (n = 60)		PROP (n = 53)		HELO (n = 57)		E2C2 (n = 7)	
DLT alone	0.74	(0.25)	0.76	(0.28)	0.78	(0.27)	0.64	(0.19)
DLT w/PMT(S)	0.81	(0.38)	0.87	(0.40)	0.95	(0.29)	0.72	(0.42)
DLT w/PMT(SR)	0.86	(0.30)	0.98	(0.35)	0.98	(0.25)	0.80	(0.24)
PMT(S) alone	3.27	(0.26)	3.42	(0.31)	3.37	(0.28)	3.41	(0.30)
PMT(S) w/DLT	2.99	(0.25)	3.14	(0.32)	3.14	(0.32)	3.15	(0.24)
PMT(SR) alone	3.58	(0.15)	3.72	(0.24)	3.69	(0.21)	3.69	(0.19)
PMT(SR) w/DLT	3.34	(0.21)	3.48	(0.23)	3.48	(0.25)	3.42	(0.16)
PMT(SRT) alone	3.66	(0.15)	3.75	(0.19)	3.76	(0.15)	3.82	(0.33)
Sgl ODCT	27.17	(13.77)	32.58	(12.01)	34.50	(13.05)	28.26	(10.73)
Sgl ADC CR	76.92	(15.74)	76.01	(16.04)	73.46	(16.47)	81.33	(13.04)
Sgl ADC IR	13.01	(26.14)	9.87	(12.06)	15.42	(39.96)	3.48	(2.29)
Sgl ADC CR/IR	26.22	(41.51)	31.29	(46.52)	18.06	(16.28)	34.75	(27.48)
Dual ODCT	31.28	(15.44)	38.76	(15.30)	39.47	(16.41)	30.80	(10.58)
Dual ADC CR	67.60	(16.97)	69.14	(16.87)	65.46	(16.31)	69.95	(13.24)
Dual ADC IR	12.67	(20.57)	9.85	(7.13)	13.75	(31.61)	4.86	(0.74)
Dual ADC CR/IR	11.48	(12.37)	15.45	(18.34)	11.53	(12.02)	14.51	(2.51)
Age (years)	22.63	(1.25)	22.69	(1.38)	23.21	(1.59)	23.14	(1.46)
Aircraft hours	74.77	(5.39)	78.08	(10.26)	78.77	(7.21)	77.44	(8.62)
Flight grade	3.08	(0.02)	3.05	(0.02)	3.02	(0.02)	3.05	(0.01)

Overall significant differences among the subgroups were found for all the PMT measures and both ODCT measures using one-way ANOVAs (all F values (3, 173) > 2.65, all p values < .05). Utilizing the Scheffe post-hoc test (p < .10), the HELO subgroup made significantly more errors than the JETS subgroup on all these measures except the PMT(S) alone while the PROP subgroup also made significantly more errors than the JETS subgroup on all these measures except the ODCT alone. See Table 7 for a complete outline of these findings. Comparing the JET and HELO subgroups to each other, as was done with the experienced pilots before, revealed significant differences on all the PMT and ODCT measures, as well as the DLT w/PMT (S) and (SR), utilizing one-way ANOVAs [all Fs(1, 115) > 3.97, all ps < .05].

The subgroups did differ significantly in terms of flight grade [$F(3, 173) = 72.21$, $p < .0001$] and aircraft hours [$F(3, 171) = 2.96$, $p < .04$] but not age. Using the Scheffe post-hoc test, the JETS subgroup had significantly (p < .10) fewer aircraft hours than either the PROP or HELO sub-groups. The

JETS subgroup also had a significantly higher flight grade than any other subgroup, while the HELO subgroup had a significantly lower grade than any other subgroup ($p < .05$).

TABLE 7. Significant ANOVAs and Intergroup Comparisons for SNA Subgroups (declining magnitude within comparisons equals improved performance).

Test measure	F	p	Significant pairwise differences
PMT(S) alone	2.85	< .04	PROP > JETS
PMT(S) w/DLT	3.40	< .02	HELO, PROP > JETS
PMT(SR) alone	4.94	< .004	HELO, PROP > JETS
PMT(SR) w/DLT	4.89	< .004	HELO, PROP > JETS
PMT(SRT) alone	5.15	< .003	E2C2, HELO, PROP > JETS
Sgl ODCT	3.50	< .02	HELO > JETS
Dual ODCT	3.59	< .02	HELO, PROP > JETS

The final primary flight grade plays an important function in pipeline assignment. For our SNA subject sample, final primary flight grade had a Pearson product-moment correlation of -.73 with pipeline assignment in which the better the subgroup performed, the lower the pipeline number. Individual pipeline choice requests and ongoing fleet requirements prevented this correlation from being higher. As previously demonstrated (11) with a larger SNA subject sample, looking only at the DLT and PMT tests, such scores are significantly correlated with this primary flight grade. Given the strong correlation between flight grade and pipeline assignment, it could be assumed that correlations between DLT/PMT scores and pipeline assignment should be significant, although analysis with this larger group remains to be done. In regards to the ODCT and ADC tests, for the smaller SNA group above, both ODCT error measures were significantly correlated (all n values > 172) with both flight grade (rs between -.16 and -.18) and pipeline (rs between .23 and .24).

CONCLUSIONS AND RECOMMENDATIONS

Our results indicate a significant difference between the performance of experienced jet and helo pilots on some of these tests. Overall, the two jet groups performed similarly, while the helo group showed a reduced level of performance in comparison. This was most evident for those PMT tests involving less complex psychomotor tasks such that, as the number of required actions for successful test completion increased, performance differences between these two types of experienced pilots decreased. These differences may have resulted partially from comparing jet pilots from fleet squadrons with helicopter pilots from training squadrons. Further research is needed to determine the possible test performance differences among other aviation communities.

Training and flight experience necessary to attain the position of fleet aviator does appear to affect one's overall test performance. This would seem evident from the large performance differences between experienced pilots and SNAs who passed primary flight training. Most of these changes were in the direction of significant improvement, but a few showed the SNAs to be outperforming one or more of the experienced groups. No definitive explanation for this occasional SNA test performance superiority is apparent at this time. Perhaps, the extra testing sessions given the SNAs on some tests improved their performance over that of other groups who did not receive this, although this was not seen for all the tests where the SNAs had such extra sessions.

Performance of SNAs who passed primary flight training differed significantly between the JET and HELO subgroups for many of the tests. Comparing the performance of these two pipeline groups with that of the experienced pilots from these same pipelines, similarities in performance differences on several tests were evident. For both the PMT(S) and the ODCT measures, whether alone or paired with a cognitive task, as well as, to a lesser extent, the DLT w/PMT(S) and (SR), differences in performance as seen for the SNAs do seem to carry through to those experienced pilots who were originally from the same SNA pipeline subgroups. Whatever psychophysiological processes are being measured by these tests, the tested ability levels of these two types of aviators appear to remain relatively constant in comparison to each other even after flight training. Significant differences on many of these tests were also seen between the JET and PROP subgroups. Further research involving the test performance of experienced nonjet fixed wing (PROP & E2C2) pilots is needed to determine whether a comparable carryover of ability levels would be found in these other aviation communities.

Besides investigating the performance differences of the nonjet fixed-wing pilot groups recommended above, an analysis of the relationship between test battery performance and other phases of training should be conducted. We do not currently know at what point in training these various tests fail to correlate significantly with flight performance. Quite possibly, these tests may only be sensitive enough to predict the relatively gross differences in flight performance seen among individuals during primary flight training. Once student pilots have been assigned to their various pipelines, the ability levels of those individuals in a pipeline may become too homogeneous for this test battery to distinguish a very good pilot from a fair one in any one particular aircraft or pipeline. Such information would prove invaluable in determining the usefulness of such tests in improving pilot selection methods and would give insight into those abilities thought necessary for a successful pilot.

This research represents a fundamental step in the development of a test battery that would reliably predict those abilities thought necessary to pilot a particular type of aircraft. Even though the current battery does not appear to predict pilot competence levels in assigned flee aircraft (7,8), at least some of these tests, especially those directly involving psychomotor functioning, could be used to predict both primary flight grade (11) and subsequent pipeline assignment. Some of these tests could even prove useful

as diagnostic tools in determining if a student pilot is performing at the level which his innate abilities, as presumably measured by these tests, would seem to indicate. A possible future area of research would involve a complete human factors task analysis of the pilot's functions in a particular aircraft, perhaps during the test and evaluation period prior to fleet introduction. This would provide a more thorough understanding of those behaviors necessary for the successful piloting of that aircraft.

REFERENCES

1. Rickus, G.M. and Berkshire, J. R., Development of an Aviation Combat Criterion: Preliminary Report, NAMI-1047, Naval Aerospace Medical Institute, Pensacola, FL, August 1968.

2. Bale, R.M., Rickus, G.M., and Ambler, R.K., Replacement Air Group Performance as a Criterion for Naval Aviation Training, NAMRL-1126, Naval Aerospace Medical Research Laboratory, Pensacola, FL, December 1973.

3. Brictson, C.A., Burger, W.J., and Gallagher, T.J., "Prediction of Pilot Performance during Initial Carrier Landing Qualifications." Aerospace Medicine, Vol. 43, pp. 483-487, 1972.

4. Shannon, R.H. and Waag, W.L., Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter Community, NAMRL-1173, Naval Aerospace Medical Research Laboratory, Pensacola, FL, December 1972.

5. Shannon, R.H. and Waag, W.L., The Prediction of Pilot Performance in the F-4 Aircraft, NAMRL-1186, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1973.

6. Griffin, G.R., Morrison, T.R., Amerson, T.L., and Hamilton, P.V., Predicting Air Combat Maneuvering (ACM) Performance: Fleet Fighter Readiness Program Grades as Performance Criteria, NAMRL-1333, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1987.

7. Griffin, G.R. and Shull, R.N., Predicting F/A-18 Fleet Replacement Squadron Performance Using an Automated Battery of Performance Based Tests, NAMRL-1354, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1990.

8. Shull, R.N. and Griffin, G.R., Predicting F-14 Air Combat Maneuvering (ACM) Performance Using an Automated Battery of Cognitive/Psychomotor Tests, NAMRL-1356, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1990.

9. Scheffe, H., "A Method for Judging All Contrasts in the Analysis of Variance." Biometrika, Vol. 40, pp. 87-104, 1953.

10. Seigel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Co., New York, NY, 1956.

11. Delaney, H.D., Validation of Dichotic Listening and Psychomotor Task Performance as Predictors of Primary Flight Training Criteria: Highlighting Relevant Statistical Issues, NAMRL-1357, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 1990.

Other Related NAMRL Publications

Dolgin, D. L. and Gibb, G.D., A Review of Personality Measurement in Aircrew Selection, NAMRL Monograph 36, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1988. (AD A200 392)

Gibb, G.D. and Dolgin, D.L., "Predicting Military Flight Training Success by a Compensatory Tracking Task." Military Psychology, Vol. 1, No. 4, pp. 235-240, October 1989.

Gibb, G.D. and Dolgin, D.L., "Validation of a Computer-based Aviation Secondary Selection System for Student Naval Aviators." Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim, CA, 1988, pp. 807-811.

Griffin, G.R., An Improved Automated Selection System For Navy Pilots, NAMRL Research Information Bulletin 88-2, Naval Aerospace Medical Research Laboratory, Pensacola, FL, June 1988. (AD A203 438)

Morrison, T.R., Complex Visual Information Processing: A Test for Predicting Navy Primary Flight Training Success, NAMRL-1338, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1988. (AD A200 394)

Nontasak, T., Dolgin, D.L. and Griffin, G.R., "Performance-based Tests, Personality Attributes, and Training Outcome Among Landing Craft Air Cushion (LCAC) Vehicle Operators." Proceedings of the Human Factors Society 33rd Annual Meeting, Denver, CO, 1989, pp. 901-904.

Shull, R. N. and Dolgin, D.L., "Personality and Flight Training Performance." Proceedings of the Human Factors Society 33rd Annual Meeting, Denver, CO, 1989, pp. 891-895.

Shull, R.N., Dolgin, D.L., and Gibb, G.D., The Relationship Between Flight Training Performance, a Risk Assessment Test, and the Jenkins Activity Survey, NAMRL-1339, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1988. (AD A200 395)